Rethinking Associations in Psychology

Abstract: I challenge the dominant understanding of what it means to say two thoughts are associated. The two views that dominate the current literature treat association as a kind of mechanism that drives sequences of thought (often implicitly so). The first, which I call reductive associationism, treats association as a kind of neural mechanism. The second treats association as a feature of the kind of psychological mechanism associative processing. Both of these views are inadequate. I argue that association should instead be seen as a highly abstract filler term, standing in for causal relations between representational states in a system. Associations, so viewed, could be implemented by many different mechanisms. I outline the role that this view gives associative models as part of a top-down characterization of psychological processes of any kind and of any complexity.

1. Introduction

Association is one of the oldest and most important concepts in psychology and philosophy of mind. In the last few years, philosophers of science have renewed interest in cleaning up the confusion that surrounds it (e.g. Buckner 2011, 2013, Mandelbaum 2015a, 2015b). But little attention has been paid to the fundamental question of what it means to say that two thoughts are associated. I describe two answers to this question in the literature, and argue that they are inadequate. I then propose a new alternative that avoids their problems.

Association is usually treated as a particular kind of mechanism that drives the sequence of representational states in a process. Many authors are explicit on this point (e.g. Anderson & Bower 1973, Beckers & Vervliet 2009, Buckner 2011, 2013, Clayton, Emery, & Dickinson 2006, Dickinson 1980, Gallo 2006, Kahneman 2011, Le Pelley 2014, Mandelbaum 2015a, 2015b, Shanks 1995, 2007), but even when the idea is implicit, it structures theory and practice. I identify two views in the literature that treat association as a kind of mechanism like this. These
two views overlap, and are often held at the same time, but it is worth separating them. Both have problems.

The first, reductive associationism, treats associative models as reducible (more or less directly) to neural network models like distributed connectionist models and neural circuit models. This implies that associative models describe a kind of mechanism in which neurons or groups of neurons are the parts, the flow of electrical activation is the primary activity, and each of these plays a representational role assigned by the model. But, as I will argue, this view conflates kinds of models that are distinct. When I discuss associative models here, I mean those in which the nodes stand in for representations in the cognitive system (e.g. localist associative models as opposed to distributed connectionist networks). Models like these are common across fields of psychology, with the Rescorla-Wagner (1972) model as perhaps the most prominent example. There are, of course, important similarities between associative models and neural network models. But they carry different claims about different features of the cognitive system. Associative models describe relations between representations, while neural network models describe relations between electrical activities in neurons. Each has a valuable role in psychology, so the models should be kept separate, and association should not be treated as a kind of neural mechanism.

The second view treats association as a kind of mechanism indirectly through the concept of associative processing. Associative processing is distinguished from many other processing kinds, including cognitive, computational, symbolic, and rule-based processing, depending on the context. I use cognitive processes, the most common contrast in comparative psychology, as the contrast here. The relevant model kinds are taken to exclude one another because they posit distinct processing kinds; distinct kinds of mechanism. Associations, then, are a feature of a
specific kind of psychological mechanism. This view produces an iterated, back and forth
dynamic between model kinds, which I illustrate using research on pigeon social learning. This
dynamic is problematic: theoretical arguments that focus on the distinction between association
and cognition are unproductive, and have become a distraction.

Both views overinterpret associative models: they take them to stronger mechanistic
commitments than they should. I argue that we should view associative models as mechanism
sketches with thin mechanistic commitments. On this view, neither association itself nor
associative processing is a kind of mechanism denoted by associative models. An association is
merely a causal relationship between representational states. ‘Association’ and ‘representation’
are so abstract that they are merely filler terms that could be realized by many different
mechanisms. This view avoids the problems with existing views: it respects differences between
different kinds of models, and it allows a more productive engagement between associative
models and other (allegedly competing) kinds of psychological model. Associative models can
be applied in many contexts in which they are not currently applied: whether you use an
associative model or any other kind depends on the question you are asking, not the process you
are describing. And associative models do not exclude other kinds: I will show how they can be
integrated with cognitive models. Associative models are simple because they abstract away
from mechanistic detail, not because they describe simple psychological processes.

I argue elsewhere that there are historical precursors for such a view of association (Dacey
2015). The goal in this paper is to develop my view of association and its opponents in detail (in
a modern context) and to show the advantages my view has. I begin by clarifying my
terminology and intended scope with respect to questions regarding both association and on
mechanisms (section 2). I then describe existing views that interpret association as a kind of
mechanism and argue that they are inadequate. In section 3, I address reductive association, and in section 4, associative processing. I present my alternative, which treats association as a filler term, in section 5. In section 6, I demonstrate how associative model should be used, especially how they should be integrated with cognitive models. I conclude in section 7.

2. Preliminaries: Associative models and Mechanistic Models

The literatures surrounding both association and mechanisms are complex. So, in this section, I explain my terminology and describe the scope and aims of my argument.

2.1. Association and Associative models

Associative models, in general, are models of psychological processes that posit associations. Associations are *links between representations that are sequentially activated in the process*. Association has been a dominant concept in thought about the mind since, at least, Hobbes in the 16th century. It was the central concept in the empiricist associationist psychologies of the 18th and 19th centuries. When that tradition fell to the behaviorists, association was reformulated but remained just as central. The cognitive revolution removed association from its place as *the* central concept, but even now it remains a core concept in the study of cognition. Today, associative models are applied frequently across psychology, used most often to describe psychological processes in nonhuman animals and human automatic processing. So association has a long complicated history and is currently used in a variety of contexts (see e.g. Gallo 2006, Mandelbaum 2015a, 2015b, reference redacted). This makes it difficult to say much about associative models in general. I intend the conclusions of this paper to apply generally, but I must limit my current discussion.
I treat association as a relation between representations in a psychological process. Some treat association as a kind of processing that does not represent. There are two ways this can go. On a (neo)behaviorist reading, association is a relation between objective events, like external stimuli and bodily movements. I take it that everything I say here could be applied to associative models so interpreted, replacing ‘representation’ with ‘stimuli’ or ‘response’ where necessary. In any case, I won’t address these uses of association in depth here; I take it that the use of association as a relation between internal representations dominates over the behaviorist use today, and it is my emphasis. Alternatively, association is sometimes treated as a relation between subrepresentational or nonrepresentational neural units, like activity in individual neurons or distributed connectionist networks. Below, I argue against the claim that these neural networks are the same as associative models. For now, I stipulate that my use of the term ‘association’ refers to a relation between representational states of the system (as in ‘localist’ rather than ‘distributed’ models). So I set both these views aside, though I should note that I am agnostic, and willing to be quite flexible, about what counts as a representation.

The term ‘association’ is ambiguous between the relation (link) between sequential representations in a process and the learning process responsible for building that relation. I take the relation between representations to be the core idea because it is included in associative learning models. Associative learning models describe the formation and modification of associations, but associations themselves are relations between representations. I’ll discuss learning when it is important, but this is how I intend the term.

I will speak of ‘associative models’ generally, but my main point is one about the nature of the association itself. With this focus, I set aside several complicating factors that are often built

---

1 The same question, whether association is merely ‘some’ causal relation between events, or a particular kind of intervening mechanism, applies to behaviorism. This can be loosely captured in the distinction between methodological behaviorism and radical behaviorism.
into associative models. Learning, just discussed, is the most common. Additionally, many associative learning models also include factors like prediction-error processes (Rescorla & Wagner 1972, Rescorla 1988), and interactions between associations (Dickinson 2001). Associative networks themselves can also become quite complex, and can be nested among processes like attention and other cognitive processes (Dickinson 2012). Factors like these might influence the amount of mechanistic detail described by the full model. I won’t argue either way here for most of these factors. Any discussion about the mechanistic content of associative models has to start with the association itself. Thus, the goal is to get the view about individual associations on the table, which can inform the interpretation of specific models in future work.

My target is the idea that association is a kind of mechanism. Views committed to this claim need not be committed to the claim that associative mechanisms must be identical. Associative processes can differ in many ways (see Rescorla 1988, Gallo 2006 ch. 1). And given the diversity of species, tasks, contexts, and levels of neural organization to which associative models are applied, from conditioning in *aplysia* to priming in humans, there is no good reason to think that these processes are united by a single mechanism (for a review of many such examples, see Papini 2008). Instead, I take each view to treat association as a kind of mechanism, which may be rather loosely cast. In the section where I discuss each view, I will make it clear what I mean by this. But before I can do so, I need to get clear on mechanistic models in general.

2.2. Mechanistic Models

The behavior of a system as a whole is a product of the organized behavior of the parts of that system. That is the idea behind the literature on mechanisms. A mechanism is a collection of...

---

I do take the diversity of systems just mentioned to make even this weaker claim suspicious, but that will not figure into my argument against it here.
parts performing activities in a coordinated way that produces the behavior of the overall system. As mechanisms are generally used, the behavior of the system is the explanandum phenomenon, and the mechanism explains how it is that the system exhibits that behavior. For instance, engines generate torque. The mechanism by which they do so includes coordinated interactions (activities) between fuel, spark plugs, pistons, and crankshafts (parts).

In general, a scientific model can be considered a mechanistic model if it describes the system in mechanistic terms. That is, if it describes the parts of the system and activities they perform (thus a mechanistic model is distinct from a model of a mechanism which is any model at all that is applied to something that is, in fact, a mechanism; see Hochstein 2016). In contrast, a purely phenomenological model does not describe the activities of parts of the system, but instead maps initial states of the system (at large) to final states (e.g. input-output mappings; Craver 2006). Phenomenological models may be sufficient for prediction and for some manipulation and control, but do not provide any details about how the system produces those input-output mappings (Glennan 2009).

In Psychology, the phenomena of interest are usually capacities or cognitive abilities (Cummins 1983, 2000); for instance, the ability to actively hold information in one’s head for a short period of time (working memory). Mechanistic explanations of these capacities break the cognitive system up into parts that perform activities that generate the relevant cognitive capacity.

Parts of a mechanism are differentiated both structurally and functionally. For instance, whether the relevant part is a neuron or a brain region including millions of neurons depends on the function the part is playing in the mechanism. Models that explicitly describe structural parts, like brain regions or distinct neural systems, are straightforwardly mechanistic models. But many
psychological models are characterized functionally, so it requires some work to uncover their mechanistic commitments. In a classic example, Baddeley & Hitch (1974) broke the capacity of working memory into three subcapacities: a central executive, a visuo-spatial sketchpad, and a phonological loop. Each of these is only described in functional terms: the central executive directs attention, the visuo-spatial sketchpad holds visual information and the phonological loop holds auditory information. However, as Piccinini & Craver (2011) note, these sub-capacities must map onto some structural decomposition of the brain if the model is to actually describe what the brain is doing, rather than being an instrumentalist prediction-generator. If the function is being performed, it is being performed by something in the brain. Thus, on any realist interpretation, functional claims like those made by Baddleley & Hitch imply structural claims, even if they are very thin.3

The formal/mathematical structure of a model does not necessitate any particular interpretation, mechanistic or otherwise (Weisberg 2013, Craver forthcoming). So any particular model could have several legitimate interpretations. Even so, I take all of the interpretations of associative models discussed here to carry some mechanistic commitments because of the basic way that associative models are used. Associative models are used in a way that treats capacities as the explanandum phenomenon. Perhaps the classic target capacity is the ability to learn certain relations between observed events: Pavlov’s dogs learned a relation between a ringing bell and presentation of food. We can explain how they learned this relation using an associative model like the Rescorla-Wagner (1972) model. So, leaving the details for later sections, associative models are mechanistic models in normal usage because they posit associations as part of an

---

3 The requirement that mechanisms be ‘localized’ in some sense is controversial (Weiskopf 2011). But for present purposes, I need not be committed to any specific relation between cognitive models and neural realizers. All that my claim requires is that the psychological model implies something about the neural realizers.
explanation of a capacity. The view I advocate here retains this basic usage, even if many of the
details change.

The explanatory role of mechanisms is controversial (e.g. Chemero & Silberstein 2008). But
for my purposes, everything I have said can be rephrased in terms of the descriptive rather than
explanatory content of the models. My question here is this: what is it that associative models
tell us about what is happening in the cognitive system when that system does what it does? This
is not a question about explanation. Even so, mechanistic terminology is helpful because it brings
together two ways of talking about differences between interpretations. I have said that
associative models are mechanism sketches, and that association is not a kind of mechanism.
These are two ways of articulating the same view, as I show now starting with the claim about
models.

Craver (2006) describes a spectrum of degrees of mechanistic commitment along which
models can be placed. At one extreme are ideally complete descriptions of a mechanism. Likely
no real models meet this standard. Mechanistic models that contain a significant amount of
mechanistic detail (but are not complete) can be called mechanism schemata. Mechanistic
models that abstract away from most mechanistic detail can be called mechanism sketches. The
different interpretations of associative models fall at different points on this continuum. The two
existing interpretations of associative models treat them as mechanism schemata (though there
differences between degree of mechanistic commitment between them), while the view I
advocate treats them as mechanism sketches. All three interpretations take associative models to

---

4 This may not be true of certain uses of behaviorist models: if association is a relation between
external stimuli (rather than representations), and one is not committed to any kind of realism about
the association itself, then one is not committed to any claims about the cognitive mechanism (see also
footnote 1). Alternatively, an association between representations could be treated as the explanandum
phenomenon (in fact, my view makes such a project more interesting than existing views; see section 6).
But this is not what associative models do.
be mechanistic models, and there is no firm line between mechanism schemata and mechanism sketches.⁵ Even so, this distinction is a useful way to summarize the concrete differences between associative model interpretations that I discuss in coming sections.

I have also said that the two existing interpretations treat association as a kind of mechanism, while the view I advocate does not. Existing interpretations take associative models to demarcate a class of associative mechanism because of the mechanistic details they take the models to carry. My view takes them to be too abstract to demarcate such a class. To clarify, my view does carry some implications about the classification of processes described by an associative model. Associative models can (in principle) be applied to any representational process. This does imply that the process represents (which is often assumed, but does not come for free; e.g. Chemero & Silberstein 2008). So, while all of the interpretations imply something about the classification of the processes they are applied to, existing views imply that there is a particular kind of associative mechanism, while the view I advocate does not.

So, the language of mechanisms is helpful in summarizing my claims about the different interpretations of associative models. But it is just that; a summary. My real arguments trade on the concrete commitments each interpretation of associative models implies, which will be discussed over the next three sections. In addition, my discussion of these concrete differences can stand or fall independent of one’s views on the role of mechanistic explanation in the cognitive sciences. With this general terminology on the table, I now discuss the three interpretations, starting with reductive associationism.

---

⁵ For instance, I say here that my view takes associative models to be mechanism sketches while associative processing implies that they are mechanism schemata. One may perfectly reasonably draw the line in a different place, such that associative processing also implies that they are mechanism sketches. I draw the line where I do to highlight the differences that matter for the current discussion.
3. **Reductive Associationism**

The first interpretation of associative models takes association to be a kind of neural mechanism. Through much of the history of the concept, it has been considered a strength that associations operate mechanically, such that they are better candidates for neural realization than more nebulous concepts like the will. In 1749, Hartley described ‘vibrations’ as the neural correlates of association (1749/1966).\(^6\) Behaviorists often held a ‘switchboard’ view where neurons simply served as relays connecting stimuli to responses (Tolman 1948). The mechanical or ‘brute-causal’ terms in which associative processes are described still motivates reductive associationism. The view found its current form and regained popularity with the rise of connectionism and of computational neuroscience in the 1980s.

Connectionism and computational neuroscience both often make use of models that describe neural systems as networks. *Neural circuit models*, a common type of model in computational neuroscience, describe the flow of electrical activity between individual neurons. *Distributed connectionist networks* model the flow of activation between subrepresentational units, which are usually thought to be functionally characterized ‘parts’ of representations. These ‘parts’ must map onto something neural for such a model to describe an actual cognitive system.\(^7\) I refer to these two kinds of model collectively as *neural network models*.

---

\(^6\) Hartley did distinguish associations between ideas from the neural vibrations that ‘cause’ them.\(^7\) Smolensky 1988. I focus on connectionist networks that are intended as models of some actual cognitive process. Many (perhaps most) connectionists do not interpret their models this way, instead treating them as *how possibly* models, or as explorations of the formalism. I set these networks aside. Note also that connectionists could build neural circuit models: my interest is in the content of the model, not the tradition in which it was built.
Neural network models are similar to associative models in several ways that motivate reductive associationism. For instance, both types of model describe sequences of cognitively relevant events using ‘activation’ spreading through a network (e.g. Churchland & Sejnowski 1992, Quartz & Sejnowksi 1997, Shanks 1994, Williams 2005, Buckner 2011, 2013, Dellarosa 1988). As a result, both types of models lack (explicit) symbolic language of thought structure (Fodor & Pylyshyn 1988). Additionally, Hebbian learning at the neural level (neurons that fire together wire together) also looks very similar to associative learning at the representation level (Clark 1993, Bechtel 1985, Bechtel & Abrahamson 1991, Walters, Carew & Kandel 1981, Le Pelley 2014). Even more specifically, prediction-error learning processes in the brain (Schultz, Dayan, & Montague 1997) look a lot like prediction-error associative processes (Rescorla 1988).

As further motivation for reductive associationism, some specific models seem to be both associative models and neural network models. For instance, in the sea slug *aplysia*, which has relatively few neurons, there are associative learning processes that can be modelled in ways such that each node stands for a stimulus or response and an individual neuron (Gluck & Thompson 1987, Walters, Carew, & Kandel 1981).

There are three general ways the reduction of associative models to neural network models could go. The strongest version identifies nodes standing in for representations with nodes standing in for neural elements, in turn identifying the representation with the neural element (or the relevant electrical activity in the neural element). For instance, debates in connectionism between proponents of distributed networks and localist networks in which nodes stand in for representations (i.e. associative models as I describe them) tend to center around the question of whether representations are localized in the brain (see Page 2000). The implication is that localist
models are illegitimate if nodes labelled as representations cannot also stand in for neural elements; the models are illegitimate if the two are not identical.

A weaker version might not require identity claims, but instead take the implementation of localist associative models in the neural networks of the brain to be trivial. The structure of associative networks and neural networks are so similar that there is no real change in the functional organization as one moves from neural networks to associative networks, as there is in the move from neural network models to cognitive, computational, or algorithmic models. Perhaps localist associative models are simply aggregates of basic neural elements and activities. For instance, Shanks says that “contemporary associative models of learning . . . are mechanistic in the sense that they react moment-to-moment to the stimulus environment and learn associative relationships on-line, and they achieve this via processes such as activation and inhibition that are known to occur in the brain” (1995, pg. 11).

Finally, one could be reductionist (in a loose sense of the term) about associative models by taking neural network models to have replaced localist associative models as the standard-bearers for associationism. The idea here is that neural network models carry forward many of the important insights of previous versions of associationism (such as the value of network structure, and the role of experience in learning). But neural network models offer significant improvements over localist associative models. For instance, neural network models can include detail from neuroscience, and they are computationally more powerful than localist associative models: there are kinds of thought that cannot be described with localist associative models, but can be with neural network models (so the thought goes, my view of association will challenge this). This view is common among connectionists. For instance, this is the view Bechtel and Abrahamsen describe when they call connectionism “associationism with an intelligent face”
I count this last view as reductionist in the loose sense that ‘higher level’ models have been replaced by ‘lower level’ models under the same name. For present purposes, though, it doesn’t matter whether each of these views count as properly reductionist by whichever notion of reduction one prefers, nor does it matter what kind of levels that notion of reduction holds between.

It can also be difficult to tell which of these views any particular researcher has. But the basic idea is the same in all of these views, and they all face the same basic problems. According to all of these views, associative models are (in some sense or other) the same as neural network models, so they directly describe structural parts of the mechanism. The parts are things like neurons, synapses, collections of neurons, brain regions, and neural pathways. Those parts perform activities like propagating electrical activation via action potentials, signaling across the synapse, and strengthening and weakening synaptic connections. Each of these neural elements then plays a representational role specified by the model (this may vary between specific models).

3.1. The Problem with Reductive Associationism

Reductive associationism treats associative models as neural models. But associative models and the two kinds of neural network model (distributed connectionist networks and neural circuit models) are each distinct kinds of model that carry different mechanistic commitments because they carry different claims about the modelled system. In associative models, nodes stand in for representational states. In distributed connectionist models, nodes stand in for structurally and functionally defined ‘parts’ of representations, and in neural circuit models, nodes stand in for individual neurons. Models of each kind can (in principle) be simultaneously applied to the same system because they are simply telling us different things about how the system works. Thus,
each of these three kinds of model carries distinct mechanistic commitments. And each can do valuable work in psychology (I’ll describe the work localist associative models do in section 6).

As I noted above, a major motivation for reductive associationism is the similarity of models of each kind. But the network architecture they all employ is a very powerful tool for describing the causal structure of systems generally (or even the non-causal structure of systems; Newman 2010). The fact that it can be applied at several different levels of organization is a feature of the formalism, not the system. The fact that learning looks similar in each is also largely a feature of the formalism. In any system that ‘learns’ (or is modified with use), learning will manifest in network formalism as strengthening or weakening links based on past patterns of activation: those are the representational resources the model has. This is not to say that there is nothing to learn from the fact that this general formalism is effective at different levels; but that fact doesn’t justify the reduction of one model to the other. Each kind of model describes entities and processes at different physical scales and differing degrees of abstraction. Each tells us different things about the cognitive system, so each answers different questions, and each is useful for different purposes. This undermines the third form of reductionism discussed above, which treats neural networks as the modern standard-bearers of associationism. Localist associative models can be their own standard bearers; neural networks models are something interestingly similar, but importantly different.

One might argue for the second form of reductionism, which claims that associative models are mere aggregates of neural networks, on the grounds that all localist associative models are (trivially) realized by an associationist neural network. I have three things to say in response to this. Firstly, there is a trivial sense in which all psychological processes are grounded in electrical activity in networks of neurons. But this does not mean that the neural architectures are
genuinely associationist in any sense that can be contrasted with other neural architectures. For instance, some neural networks genuinely compute, and some even compute digitally, despite the fact that computation is usually contrasted with associationism (Piccinini & Bahar 2013, Piccinini 2015 pg. 219). So at the neural level, networks that look superficially associationist may not be. Secondly, even if the actual neural networks are associationist, this does not mean that there aren’t levels of organization between the neural network and the associationist network at which cognitive processes could be identified (I will explain how this can be true in sections 5 and 6). So there are two ways in which this reduction could fail; even this relatively weak form of reductionism is based on a substantive empirical hypothesis. Thirdly, even if the empirical claim is true, my claim about the independent value of associative models and neural network models stands. This means that, even if we can trace the connections between the two kinds of model relatively easily, this does not justify equating them.

The strongest form of reductionism, which reduces nodes in associative models to neuronal entities and associations to neural connections, might survive the arguments so far in specific cases where the models look identical, like habituation in *aplysia*. But even then, the two interpretations (nodes as stimuli and nodes as neurons) are dissociable, both conceptually and empirically. Conceptually for the reasons I have just given. Empirically, the neural circuit model may be found to be false, but the associative model could remain true. Conflating them would lead to rejecting the associative model for the wrong reason in such a case. Even so, examples of perfect overlap are the exception, not the rule: we should not expect these strong similarities to generalize to the vastly more complicated brains of mammals, for example.

So none of the versions of reductive association are justified; the view in general mistakenly conflates distinct kinds of model. Each of the three kinds of model (associative models,
connectionist models, and neural circuit models) carries different claims about different parts of
the system, and each does valuable work. They should be kept separate, and associative models
should not be taken to carry the same mechanistic commitments as neural network models.

4. Associative Processing

The second view takes associative models to describe a kind of mechanism because it takes
them to denote a kind of psychological processing known as associative processing. The
associative processing view and reductive associationism are often held together; likely, many
take reductive associationism to imply a version of associative processing. But they are
differentiable such that one can adopt the associative processing view without reductive
associationism. In this section, I describe the associative processing view in isolation of
reductive associationism, and demonstrate its problems.

Associative processing is characterized as rigidly following the sequence of representational
states specified by associations, with associations being semantically transparent or contentless
(e.g. Mandelbaum 2015b). And associative learning is learning of contingencies between events
in the environment based on their repeated sequence. So, associative processes are often thought
to be some of the simplest processes in psychology. These are contrasted with cognitive models
that indicate more complex cognitive processes (while it is not the only contrast class, cognitive
processes will be my focus here). Cognitive processes include many posits that allow greater

---

8 Note that a reductive associationist might take (say) connectionist models to imply a kind of
processing which excludes symbolic processing, but not take this to imply that the process is simple (as
This ‘associative process’ may not face the problems discussed here, but still faces the problems of
reductive associationism.
flexibility in thought, such as cognitive maps, causal models of the world, or, in the example to come, an ability to simulate the perspective of conspecifics. The dichotomy between processing kinds makes the model kinds mutually exclusive.

Together with a preference for simpler processes, this dichotomy drives the structure of debate. The preference for simpler processes is explicitly discussed in comparative psychology, where it is traced to Morgan’s Canon (e.g. Shettleworth 2010), but the same basic idea is applied in human psychology. Since associative processes are usually the simplest on offer, the predictive success of an associative model is often taken to be sufficient to conclude that the process is associative. This automatically excludes other kinds of psychological processing.

Sometimes the inference straight from the application of an associative model to the exclusion of other kinds of models is explicit: most often when authors are pushed to defend associationist views. For instance, in one target article, Mitchell, DeHouwer, & Lovibond (2009) argue that human contingency learning produces propositional knowledge rather than associative links. In their reply, Beckers & Vervliet argue:

‘the fact that white wine makes me think of France is non-propositional (I cannot be right or wrong for being reminded of France upon smelling white wine). As such, evaluative conditioning effects very strongly appear to result from the mind being carried from one idea or representation to another, without any intermediate processing, much like what is the presumed mode of operation of an association’ (2009, pg. 201).

In another recent target article, Smith, Couchman, & Beran (2014) attack the use of associative models in comparative psychology. In response, Le Pelley argues:

‘An animal is in State X; it performs Response Y and is rewarded. When that animal finds itself in State X in the future, it will—other things being equal—perform Response Y. This is the essence of associative learning. It does not involve a ‘search’ of memory or ‘deciding’ how to respond based on inspection of memory’s contents. It is observed at the level of individual neurons, or in artificial

---

9 I discuss their arguments in section 4.3.
networks that have no notion of searches or decisions. And as such, it is very
different from metacognition’ (2014, pg. 133).

In both passages, the authors assert that an associative model (or at least, a loosely associative
description) captures the phenomenon, and take that to be sufficient to exclude other kinds of
posits. In comparative psychology, this idea has played an especially prominent role (largely
because of greater use of associative models there). To show how this influences practice in that
branch of psychology, I present an example.

4.1. Associative Processing in Comparative Psychology: Imitation in Pigeons

Because models positing simpler processes are preferred, associative models are often the
default starting point for explaining behavior. In those cases, the initial associative model is
tested by building complexities into the experimental task that the associative process could not
respond to. If the animal responds to them, the associative model is taken to be falsified in favor
of some cognitive model. In many cases, researchers will then build a new associative model that
is able to produce the observed behavior, and the process repeats itself. This sets up an iterating,
back-and-forth dynamic between associative and cognitive models.

This dynamic plays out clearly in research on imitation learning in pigeons. The simplest
imitation experiment involves two participants: a ‘demonstrator’ and an ‘observer.’ The
demonstrator is first conditioned to press a lever for a food reward. The observer is then placed
in an adjacent chamber with a view of the demonstrator and apparatus. After the observer
watches the demonstrator be rewarded upon pressing the lever, the demonstrator is removed, and
the observer is placed in the chamber with the lever. The observer is taken to have learned the
response-reward relation by observation if it presses the lever more often than controls.
At least initially, performance on this task seems like evidence that the observers process the information gained by watching the demonstrator in more complex ways than associative processing allows: associative processing does not predict learning by the observer because the observer is not actually performing the action or receiving the reward. The most common cognitive model proposed is simulation or perspective taking.

But the associative model can be made more complicated such that it predicts increased lever pressing after observation. Add a principle, known as stimulus enhancement, such that locations conspecifics have attended to are more salient (Thorpe 1963, Zentall 1996). This could compel the observer to interact with the lever. At that point, normal operant conditioning can take over. Stimulus enhancement is a relatively simple capacity, so it doesn’t violate the spirit of associative processing as standardly conceived.

Zentall, Sutton, & Shepburne (1996) tested this associative model by comparing two different demonstrator actions. In this experiment, one group of demonstrators was rewarded only when they pecked the lever, and a second group was rewarded only when they stepped on the lever. Once presented with the lever, the observers preferentially performed the action that their demonstrator had performed. Stimulus enhancement cannot explain this preference, so the authors take this to be evidence for a cognitive capacity like perspective taking.

But there is, again, a fix for associative models: add to stimulus enhancement a tendency to mimic the actions of conspecifics (mimicry – see Moore 1992, Meltzoff 1996). If an animal tends to mimic the actions of the conspecific, and does so over the lever because of stimulus enhancement, then again, standard conditioning can take over.
This process has continued to iterate. Akins & Zentall (1998) describe more complex performance that the most recent associative model can’t explain. Papineau & Heyes (2006) then describe a new associative model. Saggerson, George, & Honey (2005) then show that that model is inadequate with an even more complex experimental task. Where this leaves current theory isn’t important for present purposes. The important point is that the dynamic between these two kinds of model has the following features:

1. Associative models are the starting point, and are preferred whenever they predict behavior.
2. Associative and cognitive models are mutually exclusive: any evidence against an associative model is evidence for some cognitive model.
3. Failure of an associative model to predict behavior in a new experiment is evidence that the model is inaccurate even in explaining previous experiments.
4. As the dynamic progresses, the candidate associative model gets incrementally more complex (so long as it doesn’t get too complex to count as associative).

These features result from the assumption that association and cognition are distinct, mutually exclusive processing kinds. So even when the associative processing view is not explicit, it is an important factor in the structure of psychology.

4.2. Mechanistic Commitments of the Associative Processing View

My focus is on the associative processing view in the absence of reductive associationism. The mechanistic commitments of such a view are different from reductive associationism: it does not conflate model kinds, and it treats association as a relation between representations. Nonetheless, it sets strong constraints on the way the process operates.

Consider the logic behind varying tasks in psychological experiments. In the pigeon imitation example above, each new experiment was taken to falsify the associative model that

10 This experiment actually used the related Japanese quail, not pigeons. But this is an integrated literature, largely driven by the same labs. They see the two species as related enough to bear on one another, so I will treat them that way as well.
explained prior experiments. But each experiment used a (slightly) different task. The reason all of these experiments are thought to bear on the truth or falsity of one model is that they are thought to probe a single psychological capacity. Delimiting a capacity this way arguably introduces mechanistic commitments (as in the discussion of Baddeley & Hitch [1974] above). But more importantly, this implies that there must be some associative model that captures the activity of that process in every condition in which it drives behavior. This is why, generally, associative processes are thought to be inflexible: an associative process can only behave in ways describable by an associative model. Associative models themselves describe simple, inflexible dynamics. This severely limits the responses possible across conditions. The resulting inflexibility, in turn, is what defines the kind of mechanism known as associative processing, as I introduced it above. And because this view takes associative models to denote associative processing, any appeal to association is an appeal to the mechanism of associative processing.

This view carries weaker commitments than reductive associationism because it implies no claims about how associative processes are neurally realized. Nonetheless, it does treat association as a kind of mechanism: association is a key feature of the kind of mechanism known as associative processing.

4.3. The Problem with Associative Processing

Associative processing is unproductive. The dynamic between associative and cognitive models described above has become an impediment, rather than a virtue. Each iteration of the back and forth between associative and cognitive models produces a more complex associative model (if only marginally so). As associative models become increasingly complex, the difference between associative and cognitive processing is lessened, and the distinction is
eventually blurred. Recently, this has led authors in comparative psychology to question whether the distinction is meaningful or productive. These authors raise four related concerns.

Allen (2006) questions whether there is any single way of characterizing the distinction that applies generally across the broad range of tasks and species to which it is applied. Penn & Povinelli (2007) point out that there are many ways a process might be considered to be complex. This produces a multidimensional space that won’t easily be carved by a single distinction. Papineau & Heyes (2006) argue that some associative models are so complex themselves that the difference looks like a continuum rather than a dichotomy. Though these are similar concerns, it is telling that they worry that the distinction fails in the three different ways: failure to generalize across cases, failure to handle multidimensionality in single cases, and forcing a dichotomy on a continuum.

In addition, Smith, Couchman, & Beran (2014 and response to commentary) argue that fixation on the single question ‘is this associative or is it cognitive’ glosses over the real interesting and important psychological questions: simply stating that a process is associative tells you nothing about what the task is and requires, what information is processed, or how it is represented. They contrast the version of comparative psychology that centers on the ‘associative or cognitive’ question with a more recently developing version in which scientists ask different questions about all of the potential ways two processes can vary, and are much more flexible in their approaches to answering them. The latter, they argue, is a much more productive approach.

The associative processing view requires that we ask at what point a process crosses the line of complexity from ‘associative’ to ‘cognitive. But there are no meaningful criteria for answering
this question, and it’s not clear what we gain by doing so. Debates about whether a process, like that responsible for pigeon imitation, is ‘associative’ or ‘cognitive’ are distractions.\footnote{I do not argue that there are no processes that fall in the space traditionally called ‘associative processing’ (as does, most famously, Gallistel 1990, 2000). That is to say, there may be processes for which no psychological model can be provided except for an associative model. But this discussion does imply that we should rethink the ways we engage with processes like this. The successful use of an associative model is not sufficient reason to consider a process to belong to this class. And these processes have no special claim to associative models; rather, they are defined by the failure of other kinds of models like cognitive models. A name that better reflects the actual basis of the class, like ‘non-cognitive processing,’ would lead to less confusion.}

5. Association as a Filler Term

Both views that treat association as a kind of mechanism are inadequate. As is generally true of models, nothing forces either interpretation (Weisberg 2013, see also Craver, forthcoming). The interpretations can’t stand on their own merits, so they should be replaced. Perhaps part of the reason they have not been replaced is that there is not a systematic alternative in the literature. In this section and the next, I present an alternative and describe how it can change psychological practice.

The view I propose treats associative models as carrying the fewest mechanistic commitments of the views discussed here, treating them as mechanism sketches. On this view, any particular associative model simply describes the sequence of states a process moves through (or the influence some set of variables has on learning) at a very abstract level.\footnote{In effect, this is the inverse move of reductive associationism: Reductive associationists believe that association should be pitched at a lower level of description than it was traditionally (perhaps the level of ‘functional architecture,’ which Pylyshyn [1984] places below the algorithmic level). I argue that associative models should be pitched at a higher level of description (more like Pylyshyn’s ‘semantic’ level, similar to Marr’s [1982] computational level, though not exactly the same).} Representations are states of the system that are characterized purely extensionally, by their content. Associations are merely causal relations between those states. There are almost no commitments about how
those are realized in the brain. Association is a *filler term* for whatever causal mechanism mediates the sequence of representational states.

This interpretation does not include a commitment to the expectation that an associative model should predict behavior across conditions. Rather, an associative model should be relativized to the condition for which it was built. On this view, an associative model is properly applied if it accurately describes the way the system responds to *that* task and context. So, for instance, suppose an associative model accurately describes how a pigeon learns by observation in one of the specific experiments described above. This doesn’t necessitate that the same associative model describe performance in other, similar tasks. Of course, responses to other similar conditions can *inform* the evaluation of a specific associative model, but they don’t falsify them directly (I’ll discuss this more in section 6). So interpreted, associative models do not carry the same mechanistic commitments as the associative processing view implies they do.

Associative models do, on this view, carry some mechanistic commitments. As noted, they imply that the process is representational. In a thin sense, associative models also describe the causal structure of the capacity by describing the causal sequence of states (rather than mere input-output mappings; Craver 2006). And perhaps, as Piccinini & Craver (2011) argue, representational states should be treated as states of parts of the system, even if there are no particular commitments about what that part may be, or what that part is doing.

This view avoids the problems with the other views. It keeps associative models clearly distinct from neural network models. Though a neural network model might (or might not) describe the mechanism underlying an association, the association should not be reduced to that neural network. It also eliminates the mutual exclusivity of associative models and cognitive models, avoiding the dynamic that makes associative processing unproductive. Because
associative models are abstract and limited in scope, they do not exclude other kinds of model from being applied to explain the behavior of a capacity that has already been described by a cognitive model.

Of course, if two models make different predictions in a specific condition, they exclude one another. This will often be true of candidate associative and cognitive models. But this is a very different issue than the blanket exclusion associative processing implies; this is interplay between particular models, not by the kinds to which two models belong. The fact that a cognitive model makes different predictions across conditions does not mean it conflicts with the associative model if they make the same predictions within a condition. These other kinds of models may provide more mechanistic detail, or may predict behavior across a larger range of conditions. An association could be underwritten by the application of a rule in one step of an algorithm, manipulation of a cognitive representation like a mental map, or, by any kind of neural network (including those that are digital computers; Piccinini 2015, pg. 219).

Associative learning models also describe the influence some set of environmental variables has on learning. Traditionally, these are limited to simple variables like patterns of co-occurrence. But patterns of co-occurrence are surely important variables in many kinds of learning, so co-occurrence based learning models can be predictive even if they do not completely characterize the process. In other words, if co-occurrence is one of the main variables that a system exploits in learning, then a model tracking co-occurrence will get predictions right much of the time. This does not mean that co-occurrence is the only variable the system tracks, as the concept of associative processing is often taken to imply. Indeed, the experimental paradigms in which associative learning is usually tested intentionally exclude other variables. So the fact that models that only include co-occurrence are predictive in these cases is just as
likely a result of restrictions in the experimental environment as it is restrictions in the flexibility of the learning system. There is no principled reason to think these systems are restricted to co-occurrence, along with a few other simple variables. And on this view there is no principled reason to limit the learning variables included in the model.

In general, whether you use an associative model or any other kind depends on the question you are asking, not the process you are describing. This gives associative models a significantly different role in psychology.

6. Integrating Associative Models and Cognitive Models

The fact that associative models are compatible with other kinds of model means there should be a much more complex, constructive interplay between associative models and cognitive models than current practice allows. The associative processing view treats associative models as the starting points to inquiry because it takes them to describe the simplest processes, which are preferred. Treating association as a filler term takes associative models to be a starting point because they abstract away from mechanistic detail. This means that they are useful when we don’t know enough about the mechanistic detail. They can be an early step in a top-down characterization of the process (that is, moving from and abstract, partial characterization of the process to one including more causal and mechanistic detail). In this section, I sketch the general approach that integrates associative and cognitive models as part of this top-down project.

To do so, I return to the pigeon imitation learning example. The two candidate models discussed above are an associative model and a perspective taking model. The perspective taking model is not very well filled out, and associative models, on my interpretation, can help elaborate it. I demonstrate by proposing three different variants of the perspective taking model, along with
two alternative models that do not require a perspective taking capacity. I design these to predict performance in the experiment described above in which the observer pigeon had to press the lever by the same means the demonstrator did to receive a reward, either by pecking or stepping (Zentall, Sutton, & Shepburn 1996). For simplicity, I’ll just discuss cases where the demonstrator pecked the lever: for successful performance in this condition, the observer has to be somehow induced to peck a lever by observing a conspecific receive a reward for doing so. This is what the model must predict.

In all of these proposed models, words stand in for representational states (specifying them by content), and arrows stand in for associations. I use square brackets to indicate those states occurring within a simulation, italics to indicate a representation of a stimulus, and regular text to indicate a representation of a response. Causal links between events in the word that bring about representations are not included in the models, for simplicity.

On the most basic interpretation of the perspective taking model, the observer could simulate the global perspective of the demonstrator, and then effectively undergo classical conditioning within the simulation:

1a. \[\text{peck lever} \rightarrow \text{reward}\]

The relation is learned within the simulation, resulting in:

1b. \text{peck lever} \rightarrow \text{reward}

On my interpretation of associative models, this is an associative model both of the capacity to simulate and of the relation learned by doing so. One might reply that this associative model does not get directly at what is really interesting about the perspective taking capacity, which is the ability to simulate. But if the basic structure of this model is right, we need to figure out what
sequences can occur within the simulation and how it is triggered in order to characterize the capacity. This is exactly the role in which associative models are helpful.

Such a global simulation might seem unlikely, since it would seem to imply that pigeons are constantly simulating the perspective of conspecifics. An alternative might be a more piecemeal simulation, in which certain actions by the conspecific trigger simulations for specific, limited, periods of time. For instance:

2a. conspecific pecks lever → [peck lever]

Followed immediately by:

2b. conspecific receives reward → [reward]

The observed actions of the conspecific can then drop out, and the temporally paired simulated representations can again form:

2c. peck lever → reward

Alternatively, the simulation could be triggered by the reward, prompting the retrieval of previous actions made by the demonstrator from short term memory in the form of a simulation:

3a. conspecific receives reward → [peck lever → reward]

3b. peck lever → reward

These possibilities are far from exhaustive, but they are instructive: there are many ways the perspective taking model can operate, and associative models can formalize the differences between them.

Compare these to models that do not involve simulation. The associative model that was in fact proposed to explain behavior in this experiment added stimulus enhancement and mimicry to standard classical conditioning (Moore 1992, Meltzoff 1996, Akins & Zentall 1998). According to this model, during the observation period the observer simply gains a prepotent tendency to
pay attention to the lever, and a similar prepotent tendency to peck. The reward serves as a general motivator to act. I’ll represent these motivations with curly brackets.

4a. Conspecific pecks lever → {peck}, {lever}

Then:

4b. Conspecific receives reward → {increased general motivation}

Then, during the test period:

4c. {increased general motivation, peck, lever} → peck lever

And upon pecking the lever:

4d. reward

Producing:

4f. peck lever → reward

Aside from the lack of perspective taking, this model differs from the models above in that the relation between lever pecking and rewards is not learned during observation. Instead, observation motivates behaviors that lead to fast learning during test. We could produce a model that lacks perspective taking, but still describes learning during observation. Consider the model below. First, during observation, the pigeon observes the relation between a conspecific pecking and receiving a reward:

5a. Conspecific pecks lever → conspecific receives reward

Then, through some general abstraction process that doesn’t involve any actual simulation:

5b. peck lever → reward

Here, learning occurs during observation without simulation.

These are all models describing the sequence of representations in the process responsible, so they are all associative models. Perhaps the first two, which include cognitive elements as
well as associations, could be considered ‘mixed models,’ though the associative elements are crucial. The associations imply nothing more about the process than that the representational states go in the causal order specified. Each of these is a very abstract characterization of the process responsible; none of these implies that the process described is associative. Even (4), which was a straightforward transcription of the associative model in the literature, might describe one manifestation of a complex mechanism that manifests differently in other contexts.

They all predict the behavior in this experimental condition, so this test doesn’t motivate any one over the others. Of course, more complex conditions have been tested, and the birds were able to learn by observation in those conditions (Akins & Zentall 1998, Saggerson, George, & Honey 2005). Those tests don’t themselves falsify any of these models, because each model only needs to apply to the condition for which it was built. The way that different experiments bear on one another is more complex. First, one would build a similar set of possible associative models that predict behavior in the new experiment. Then, adding in the assumption that there is a shared capacity, attempt to determine which sets of associative models that predict in each experiment might all be manifestations of a single mechanism. Of course, if one can apply the same associative model to a capacity across contexts, that is an important result. Or if there are structural similarities between the predictive associative models, that general pattern provides important information about the mechanism. But even this isn’t necessary; a single mechanism may produce very different associative sequences in different conditions. In such a case, inferences from the associative models to the nature of the process may be difficult, but the constraints set by each associative model on the mechanism are just as strong in any case.

One should not reject an associative model of performance in some condition because it fails to predict in a different condition; but one should reject an associative model that could not be
the product of the mechanism that best explains behavior across conditions. So there are two steps here. First, one builds candidate associative models that predict results in particular contexts, then one compares those models and attempts to characterize a mechanism that could produce (at least) one of the candidate associative models in each context. This process might be quite complex, perhaps taking into account theoretical virtues like unification and simplicity. It is only at the second step, when one looks beyond the associative models themselves, that one implicates any particular kind of mechanism.

As presented here, these models are quite abstract, even by associative modelling standards. One could mathematically describe the dynamics of learning, perhaps using variants of the prominent Rescorla-Wagner (1972) model. This would add more detail to the model, perhaps increasing mechanistic commitments somewhat. But there additional constraints do not mean that the models should be taken to describe a kind of mechanism, since the arguments above still apply. Following the interpretation advanced here, they still describe the influence that repeated pairings have on subsequent sequences of representations in the mind.

All of the models just described were also built ad-hoc. That doesn’t mean that associative models in general must be. But even if they usually are, it is not pernicious. The worry about ad-hoc models is that they simply tack a new epicycle on as each new experimental result comes in. This is not the project I have described. There should be an ongoing attempt to integrate associative models using a model of a mechanism that can produce them.

This discussion has applied associative models to a complicated variant of classical conditioning, one of the paradigmatic domains of association. There likely are experimental

---

13 The Rescorla-Wagner model describes a process whereby associative strengths are adjusted based on errors in predictions that association generates. But, putting the point above more specifically, this should not be taken to imply a kind of mechanism of prediction error. That would run into my arguments against associative processing. In general, most in the field do not take the ‘prediction’ literally or realistically, even if they do so with the association (Danks 2013, 2014).
paradigms for which associative models are either easier to apply, because the experiments gain more direct access to the sequences of states in a process, or more helpful, because the sequences explain the phenomena more directly. Most of the paradigmatic associative phenomena have one or the other property: associative learning, priming, and word association task responses are all phenomena in which the sequence of representational states is somehow central. This does not mean that associative models should be restricted to experiments like these (which would effectively replace the theoretical restrictions in the use of associative models that I have rejected with empirical restrictions). It is simply a matter of getting creative with associative models, in ways similar to those just described.

7. Conclusion

Current use of associative models is dominated by interpretations that take them to describe a particular kind of mechanism. These views are inadequate. Reductive associationism conflates distinct model types that have independent value, and associative processing leads to unproductive practice. If we treat association as a filler term, these problems are solved. Such a view of association does not conflate relations between representations with relations between neurons, as does reductive associationism. It does not pit associative models against cognitive models in a zero-sum game, as does associative processing.

This view of association gives associative models a new role in psychology, in which they set top-down constraints on the mechanism. If some capacity involves a certain sequence of representational states, then the mechanism must capture that. At each step, the comparison is between specific models, not model kinds. This unproblematically allows mixed models including associative and cognitive elements, and associative models of cognitive processes. Indeed, as shown, associative models are genuinely helpful in characterizing cognitive processes.
This is an important point: it is often the case, as it is in the pigeon imitation example, that
cognitive models are not rigorously or mathematically characterized. Associative models, in
contrast, are popular in large part because of their mathematical rigor. There are many reasons
why it is very hard to provide equally rigorous cognitive models. Associative models, properly
interpreted, can help.

References


and Sons.

learning and motivation*, 8, 47-89.

models cognitive or associationist?. *Behaviorism*, 13: 53-62.

Parallel Processing in Networks*. Blackwell. Cambridge, MA.

learning. *Behavioral and Brain Sciences*. 32, 200-201

Buckner, C. (2011). Two Approaches to the Distinction between Cognition and “Mere
Association.” *International Journal of Comparative Psychology*. 


